



**A PROPOSED GUIDELINE
FOR
THE EXCHANGE OF IMAGES AMONG
CENDI AGENCIES**

Submitted by
**Image Exchange Focus Group
CENDI Information Exchange Working Group**



Prepared by
**Gail Hodge, CENDI Secretariat
Information International Associates, Inc.
Oak Ridge, Tennessee**

April 1997

CENDI IMAGE EXCHANGE FOCUS GROUP

Chair	- Barbara Bauldock (DOE/OSTI)
DOE/OSTI	- Jannean Elliott*, Ken Hohenbrink, Norm Smith
DTIC	- Helen Viel*, Susan Ruddle
NASA	- Roland Ridgeway*, Roy Stiltner, Patricia Baxter
NLM	- Lou Knecht*
NTIS	- Tom Pennington*, Roger Counts, Sue Feindt

* Indicates primary contact.

CENDI is an interagency cooperative organization composed of the scientific and technical information (STI) managers from the Departments of Commerce, Energy, Defense, Health and Human Services, Interior, and the National Aeronautics and Space Administration (NASA).

CENDI's mission is to help improve the productivity of Federal science- and technology-based programs through the development and management of effective scientific and technical information support systems. In fulfilling its mission, CENDI member agencies play an important role in helping to strengthen U.S. competitiveness and address science- and technology-based national priorities.

Table of Contents

EXECUTIVE SUMMARY.....	1
1.0 INTRODUCTION.....	2
2.0 IMAGE FORMAT.....	2
3.0 DIRECTORY AND FILE STRUCTURE.....	2
4.0 CONNECTION TO BIBLIOGRAPHIC RECORDS.....	3
5.0 DIRECTORY FILE.....	3
6.0 DISTRIBUTION MEDIA.....	4
7.0 QUALITY CONTROL.....	4
8.0 MAINTENANCE OF GUIDELINES.....	4

EXECUTIVE SUMMARY

This document provides guidelines for the exchange of images among the CENDI partners. It is meant to provide assistance for project and technical staff seeking to import or export images. As the CENDI agencies move toward electronic acquisition and distribution of documents, the continued exchange of hardcopy and microfiche documents does not fit into the electronic workflows envisioned by the agencies. In order to promote the efficient sharing of images, the CENDI Principals (July 16, 1996) requested that the Information Exchange Working Group develop guidelines for sharing images.

The standard format is compliant with the TIFF 6.0 Specification with CCITT Group 4 compression. A TAR file is created for each individual document and separated from the next TAR file on the tape by an EOF mark. Each page of a document is stored as a separate image in a separate file. The filename corresponds to the page number or the order of the images within the document. The filenames are eight digits in length, padded to the left with zeroes. Based on this convention, the filenames can be computed easily, eliminating the need to store the filenames elsewhere in the system, and providing an easy way to ensure proper presentation order of the images. The filename extension is ".tif", serving as an indication of the viewer needed to display the file. For example, the filenames are 00000001.tif, 00000002.tif, etc. The filenames for a document are stored in a directory named for the document ID assigned by the originating agency. In this way, the document can be manipulated by the directory name rather than the individual filenames. The unique document ID is also the link to the bibliographic record, which records, if provided, are stored on a corresponding tape.

The preferred distribution medium is 8mm DAT. The tape includes an internal directory file, external label, and transmittal report that provide information regarding the filenames within each directory, the number of image files and documents, the sequence number of the tape, the total number of tapes being sent, the originating agency, and the date or volume and issue of the announcement.

The minimum resolution for scanned images is 300 dpi. Higher resolution may be used as necessary to achieve high quality images or for specific image types. Care should be taken to provide the best quality available from the original.

These guidelines were developed February 1997. They may be modified as the technology changes and the sharing of images between CENDI agencies increases production-level experience.

1.0 INTRODUCTION

This document provides guidelines for the exchange of images between the CENDI partners. It is meant to provide assistance for project and technical staff seeking to import or export images.

Many of the CENDI agencies have historically shared hardcopy and microfiche formats of documents in order to reduce the redundancy of acquiring, cataloging, and in some cases indexing documents covered by one agency that are also relevant to another CENDI agency's scope and coverage. As the CENDI agencies move toward electronic acquisition and distribution of documents, the continued sharing of hardcopy and microfiche documents does not fit into the electronic workflows envisioned by the agencies. In order to promote the efficient sharing of images, the CENDI Principals (July 16, 1996) requested that the Information Exchange Working Group develop guidelines for sharing images.

A focus group meeting with representatives from the Defense Technical Information Center (DTIC), the Department of Energy Office of Scientific and Technical Information (DOE OSTI) (Washington), the NASA Center for AeroSpace Information (NASA CASI), the National Technical Information Service (NTIS), and the National Library of Medicine (NLM) was held on July 24, 1996. Representatives from DOE OSTI in Oak Ridge, TN participated via videoteleconference.

2.0 IMAGE FORMAT

The standard format is compliant to the TIFF 6.0 Specification with CCITT Group 4 compression.

3.0 DIRECTORY AND FILE STRUCTURE

A TAR file is created for each individual document and separated from the next TAR file on the tape by an EOF mark. Do not use any "hard path" for the document directory.

The directory name is the unique document ID assigned by the originating agency. Segregating the documents by directory facilitates moving the document to a new disk or copying it to tape, since the manipulation is by directory rather than by individual filenames.

The files containing the TIFF images for each page of the document are stored within the directory. The naming scheme for the image files is the image number within the document followed by the ".tif" extension. The image number is the sequence of the images in the original document, padded to the left with zeroes to create an eight-digit numeric. For example, a five-page document has the image filenames 00000001.tif, 00000002.tif, 00000003.tif, 00000004.tif, and 00000005.tif. Gaps in the sequence of filename numbers, such as those caused by blank pages in the original document, are permissible as long as a numerical sort of filenames results in the images falling into the correct sequence.

This naming convention is compatible with most computer systems. (Note that this convention is not directly compatible with MS-DOS systems if the document ID, i.e., the directory name, exceeds eight characters.) In addition, the filenames are computed easily, eliminating the need to store the filenames elsewhere in the system. The filenames also indicate the order of presentation.

For example, create the tar files like:

```
tar cvf /dev/nrst0 documentID1_dir
tar cvf /dev/nrst0 documentID2_dir
.
.           [where documentID1 and documentID2 are the unique ID's for the documents
.           assigned by the distributing agency]
.
```

The listing resulting for tar tvf /dev/nrst0 is:

```
/documentID1_dir/00000001.tif
/documentID1_dir/00000002.tif
.
.
/documentID1_dir/end.tif [where end is the last sequential image number in documentID1]
EOF
/documentID2_dir/00000001.tif
/documentID2_dir/00000002.tif
.
.
/documentID2_dir/end.tif [where end is the last sequential image number in documentID2]
EOF
```

4.0 CONNECTION TO BIBLIOGRAPHIC RECORDS

For exchange of new records, the full bibliographic records are included on a separate tape. (This Guideline does not include the specification for the bibliographic record.) The bibliographic record has the same unique document ID used to name the image file directory so that the two records can be connected by the receiving agency.

The transfer of TIFF images for archival documents may be accompanied by bibliographic records, depending on the agreement between the agencies. If bibliographic records are included, they may contain fewer elements than the full record. (These specifications will be more clearly defined following future discussion of legacy collections.) If images are being sent for bibliographic records provided previously (i.e., the scanning of legacy documents), these images are provided on a separate, specifically marked tape.

5.0 DIRECTORY FILE

Each tape includes a file containing the directory name and the filenames within each directory. The name of this file is tape_toc.

6.0 DISTRIBUTION MEDIA

The preferred distribution medium is 8mm DAT. The external label minimally identifies the originating agency, the type or classification of data included, the date range (or announcement volume and issue) for the selected records, the sequence of the tape and total number of tapes in the set, and the date the tape was created.

A transmittal report or packing slip accompanies the tape on paper or as an electronic file. The report minimally includes the agency or database name, the date range or the announcement volume and issue numbers for the documents selected, the date the file was created, the number of files, the number of documents, the number of pages (images), and the list of the filenames for each directory.

7.0 QUALITY CONTROL

The minimum resolution for scanned images is 300 dpi. Higher resolution may be used as necessary to achieve high quality images or for specific image types. While there are no quality standards for images, care should be taken to provide the best quality available from the original. Some guidelines include:

- Beware of “stray images” resulting from scanning on a flatbed scanner;
- The document must be viewable by a standard TIFF viewer that supports TIFF Group IV;
- 400-600 dpi resolution is recommended for color and halftone photographs.

8.0 MAINTENANCE OF GUIDELINES

These guidelines may be modified as the technology changes and the sharing of images between CENDI agencies provide the CENDI agencies with production-level experience. The CENDI Information Exchange Working Group is responsible for periodic review and update of these guidelines.